Problems Associated with Collecting Drinking Water Quality Data for Community Studies: A Case Example, Fresno County, California

M. DONALD WHORTON, MD, ROBERT W. MORGAN, MD, OTTO WONG, ScD, SUZANNE LARSON, MPH, AND NANCY GORDON, MS

Abstract: This paper discusses methodology in developing exposure data for the water supply contaminant dibromocholoropropane (DBCP) in Fresno County, California. There are 532 drinking water systems (49 large and 483 small) within Fresno County plus 14,000 private wells. We determined the number of wells per system, the output per well, and the population served by each system. The task of deriving water quality estimates for each census tract was complicated by the fact that a single census tract can be served by more than one system; each system usually has more than one well; and a single well can have several episodes of testing for various contaminants. We calculated a series of weighted averages for concentrations of DBCP, arsenic, and nitrates for each census tract, using water production figures for each well as the weighting factor. Water quality data were derived from a total of 14,861 laboratory reports, although the majority did not report on all contaminants. Mean DBCP levels ranged from 0.0041 ppb to 5.7543 ppb among the census tracts. We found no correlation between DBCP levels per census tract compared to either arsenic or nitrates. We believe that we made as complete an exposure assessment as practically feasible. (*Am J Public Health* 1988; 78:47–51.)

Background

At this time, there is considerable concern over the safety of drinking water in the United States. Chemical rather than bacterial contamination now stimulates public concern, epidemiologic studies, regulation, and litigation. In 1986, Californians expressed their collective concern about the effects of chemical carcinogens and reproductive toxicants by the passage of Proposition 65. Epidemiologic studies of the effects of chronic exposure to contaminated drinking water are difficult due to lack of direct exposure data, the multiple sources of drinking water, population migration, possible confounding by several factors, and probable low dosages of contaminants.

In 1984, we began a series of studies investigating the relation between drinking water exposures to the nematocide dibromochloropropane (DBCP) among residents of Fresno County, California, and specific health outcomes. The genesis of the studies was the finding of widespread DBCP contamination of the drinking water wells in various farming areas of California, the unpublished report by the California Department of Health Services indicating an association between DBCP contamination of community drinking water in Fresno County with both stomach cancer and lymphatic leukemia,[1] and the well recognized effects on fertility of male workers exposed to DBCP.[2,3]

This article describes the problems encountered and the assumptions that had to be made in defining a population served by an identifiable water source in an area with multiple water sources. The primary contaminant variables of interest were DBCP, arsenic, and nitrates.

Drinking Water Sources

In Fresno County, most drinking water, other than commercial bottled water, originates from groundwater pumped and distributed by a number of purveyors with systems of varying size. The county has three categories of water systems: large systems (200 or more service connections), small systems (5-199 connections), and private wells serving one to four residences. For large systems, the state has responsibility for water quality; for small systems the state has delegated that responsibility to the Fresno County Health Department. Since 1977, the County has also regulated private wells.

Large Water Systems

The State Sanitary Engineering Regional Office in Fresno keeps records of water quality test results for each large system by specific well. Routine water quality testing for each well is done triennially. Each water quality test has a separate report which includes water system identification, sampling data, well number, location, water constituent or contaminant tested, the specific test results, measurement unit, and name of testing laboratory. Routine water quality samples must be done by state approved laboratories. Quality control for laboratory results is one of the criteria for state approval.

The routine water quality testing includes general mineral and physical analyses, inorganic chemical analyses, organic chemical analyses (chlorinated hydrocarbon pesticides, chlorophenoxy herbicides, and trihazomethanes) and radioactivity, both natural and man-made. Maximum Allowable Contaminant Levels (MCLs) are the legally enforceable drinking water contaminant levels set by the US Environmental Protection Agency and adopted by the State of California. MCLs are not established for all chemicals tested. Not all water quality reports have data on all the different pesticides and organic solvents listed since concerns about these have been an evolving process. All of the water quality test reports from 1960 through 1984 for each water system and each individual well were on hard-copy; they had not been computerized.

The State began collecting DBCP information in June of 1979 as a special project, and these results were kept in

From Environmental Health Associates, Inc. Address reprint requests to M. Donald Whorton, MD, Environmental Health Associates, Inc., 520 Third Street, Suite 208, Oakland, CA 94607. This paper, submitted to the Journal September 29, 1986, was revised and accepted for publication June 22, 1987.

notebooks separate from routine water quality data; the notebooks extended from June 1979 through 1984. In addition to DBCP levels for large water systems, the state notebooks contained results from some small systems and some private wells. The Department of Health Services Sanitation Laboratory in Berkeley conducted the analysis of DBCP in the water. The California Department of Food and Agriculture had also collected some DBCP well data in 1979.

We developed and sent a questionnaire to each large water system manager to obtain descriptive information about the system. Questions included: name of the system, boundaries of the service area, type of water sources, number of wells, type of well system (open or interconnected versus closed or not interconnected), enactment and abandonment date for wells, and well production for the years 1960–83. We followed-up by telephone when completed questionnaires were not returned. Of the 49 large system managers, 42 cooperated by sending us the necessary information or allowing our staff to visit their offices and collect the water systems information. We obtained limited data on the seven remaining large systems from State and County records. This data limitation did not diminish, to any substantial degree, our large system data.

We developed a different questionnaire tailored for the small systems to their respective managers, with follow-up similar to that described for large water systems. We obtained completed questionnaires from only 130 of the 483 systems, and for the remainder, we used the limited data obtained from the County. Because of the smaller size of these systems, the County records were adequate substitutes for the completed questionnaires.

Small Water Systems

The Fresno County Sanitary Engineering Office keeps hard-copy records of water quality test results for small water systems. Most small systems have a single well. As with the state supervised systems, each well is required to be tested triennially. The water constituents reported and tested are similar to those previously described for large water systems.

Private Wells

Some private well data were available from the State. The County would not allow access to the individual private well data if any individual identifiers were included. This confidentiality was necessary to protect property values and to inhibit sales solicitation for water purifying attachments. We identified the well as being located within one square mile as defined by the geological survey identifiers of Township, Range and Section (TRS). All of the well data kept by the State were identified through TRS, and each TRS was treated, in subsequent analyses, as a water system.

There are three separate data sets for private wells from the County:

- a study initiated by the County on DBCP levels in 1,000 wells;
- approximately 1,500 well samples in which the well owners requested the county to test for DBCP;
- water quality analyses for all new well permits issued since 1977.

To facilitate access to the new permit data, we offered to computerize these data for the County, free of charge. We kept our agreement regarding confidentiality for the data we utilized.

Analytic Needs

For different types of epidemiologic studies, we needed different types of information concerning drinking water distribution. For population or ecologic studies, we needed to relate water data to population denominators; we utilized US Census Bureau data that were available by census tract. However, since many census tracts had more than one water system, we had to develop an exposure estimate for each census tract. For case-control studies, we needed to determine the drinking water source and quality for the residence of the individual. Thus, we needed to determine a value for the water system that served that residence.

Determining Contaminant Values

We first calculated the mean contaminant value for individual wells within each of the large water systems. For wells with the results of less than detectable, we used one-half the detection level in all calculations. In order to calculate a contaminant level for the specific water systems, we weighted each well by annual production volumes. Thus, for each large water system, a contaminant value was determined by this mean weighting system so that each well was proportionally represented.

The contaminant levels of small water systems were much simpler to calculate because few small systems had more than one well. For those with more than one well, we used the same process as for the large systems. For systems with only one well, we took the mean of all applicable results.

Private well data were analyzed by TRS. We took the mean contaminant value for each TRS and treated each TRS as a private well system in a manner similar to our treatment of small and large water systems that had more than one well. Because we did not have individual identifiers for each private well, we could not determine if any individual well had more than a single measurement.

Demographic Information

Table 1 shows the descriptive characteristics of the 49 large and 483 small drinking water systems in Fresno County. One of the large systems uses both well water and surface water. Well water is the only source for 46 large systems, while two systems use only surface water.

There are 397 wells in all of the large systems, with an average of 8.0 wells per system (range 1-110). There are 594 wells in the small systems for an average of 1.2 wells per system (range 1–10).

We did not have water production data for all large or small water systems. The results for the 39 large systems with such data showed a total production of 377,999 million gallons per year with an average of 9,692 million gallons per system. For the 47 small water systems with production data, total production was 11,167 million gallons per year, averaging 238 million gallons per year per system. For water systems in which we had no production data, we utilized the State standard estimate of 3.9 persons per service connection and 250 gallons used per person per day. These numbers were used whenever we lacked water production data for contaminant level whenever we lacked water production data for contaminant level calculations, e.g., mean DBCP levels. The use of a standard estimate due to the lack of production data for each well could overestimate or underestimate the amount produced by each well, but likely represents the average estimate.

There are no accurate data regarding the number of private wells within the County although the County author-

TABLE 1—Water Systems Descriptive Statistics

	Large Systems	Small Systems
Number of Systems	49	483
Water Source		
Well only	46	322
Surface only	2	113
Both	1	40
Not stated		8
Water Production* (millions gallons/year)		
Total Systems	377,999	11,167
Average System	9,692	238
Well	1,189	123
Well Systems		
Total Number of Wells	397	594
Range # of Wells/Systems	1-110	1-10
Number of Open Systems	49	104**
Range Dates Drilled	1913–84	1920–84
Number Abandoned	51	38
Range Dates Abandoned	1958–83	1965–84
Average Number of Systems per Census Tract	1.6	9.5

*Water production figures are limited to those 39 large systems and 47 small systems for which we have production data.
**Multi well systems only.

ities estimate that approximately 10 per cent of the population uses private wells for drinking water. If we apply the generally accepted estimate of 3.9 persons per hook-up (in this case well), there are between 13,000 and 15,000 private wells in the County used for drinking water purposes. As shown in Table 2, we have some type of water quality data on approximately 7,000 private wells or about half of the estimated total.

Water Quality

Water quality descriptive information is shown in Table 2 for each of the three types of water systems. As previously stated, a TRS is considered equivalent to a water system for private wells. For the large, small, and private well systems, we had separate water quality information from 4,709 reports, 2,834 reports, and 7,318 reports respectively. Information on DBCP was available from 1979 through 1983 for large and small systems, and for 1979 through 1985 for private well systems. Nitrate data were available for large and the small systems from 1961 to 1983, and from 1963 to 1983, respectively, but were only available for private wells starting in 1970. Arsenic data are less complete, as shown in Table 2.

Although we had a large number of water quality reports (14,861), few reports contained results on all chemical contaminants. For example, we had approximately 7,000 private well test reports, but only about 4,000 of those contained DBCP data. The most commonly reported contaminant is nitrate, with 6,359 results reported from a total of 14,861 water quality tests. Thus, 43 per cent of all tests had nitrate results. There were 7,207 DBCP samples and 1,562 arsenic samples. No other pesticide or industrial chemical had more than 250 samples.

TABLE 2—Water Quality Descriptive Statistics

	Large	Small	Private Wells*
Number of Systems	49	449	1,257
Number of Test Reports	4,709	2,834	7,318
Years of Test Reports			
DBCP	1979–83	1979–83	1979–85
Nitrates	1961–83	1963–83	1970–85
Arsenic	1961, '66, '70 '73–'80	1971, '73, '77–'83	1978–83

*All wells within a TRS are defined as a system. (Each test report does not include all chemical analyses.)

Water quality test results show the respective DBCP mean values for the three system types (large, small, and private wells): 0.6 ppb (range < 0.001 to 360 ppb), 0.5 ppb (range < 0.001 to 36 ppb), and 1.4 ppb (range < 0.001 to 140 ppb). Generally, water quality data are expressed in milligrams per liter (mg/l). However, with DBCP the allowable level is 0.001 mg/l (1 μg/l) or 0.001 ppm (1.0 ppb). We have found it easier to express DBCP results in ppb or micrograms per liter (μg/l).

Nitrate results for the three system types (large, small, and private well) show the following mean results: 14.2 mg/l (range < 0.005 to 66 mg/l), 14.9 mg/l (range < 0.005 to 148 mg/l), and 15.0 mg/l (range < 0.0003 to 340 mg/l), respectively. The respective arsenic results were 0.005 mg/l (range < 0.001 to 0.25 mg/l), 0.003 mg/l (range < 0.0006 to 0.2 mg/l), and 0.016 mg/l (range < 0.005 to 0.07 mg/l).

Contaminant Levels by Census Tracts

There were two generic issues for each census tract: which water system supplied drinking water to the census tract, and within multi-systems, what representative proportion of the census tract was served by each specific water system. The average number of large water systems per census tract is 1.6; whereas the average number of small systems is 9.5.

Large water systems generally served more than one census tract. We elected to use the same mean contaminant value for a specific large water system in all census tracts that the system covered. For small systems we had the TRS for the address of the system and the well. Since we also knew the census tract of every TRS in the County, we categorized small systems by appropriate census tracts.

There were four possible combinations of types of water systems within a census tract. The possible combinations and the methods are described in the following:

1. *Census tracts with only one type of water system—*

We used the specific system's mean contaminant value as previously described for the entire census tract.

2. *Census tracts with a large water system plus either a small water system or private wells*—By use of mapping techniques, we determined the geographic percentage of the census tract supplied by the large water system. If there was more than one large water system, we determined the geographic percentage for each system. We assigned the remainder of the census tract to the other type of water system. We weighted the results for each type of system based on these geographic percentages.

3. *Census tracts with a large water system plus both one or more small water systems and private wells*—We used the same method as described above (2) for the large water systems. Since we did not have accurate boundary zones for each small water system, we could not map them in the same manner as the large systems. For the remainder of the census tracts, where we had both small water systems and private wells, we gave equal weight to each of their contaminant values.

4. *Census tracts with only small water systems and private wells*—We gave equal weight to each type of water system. If there was more than one small water system, we divided the total contribution of all the small systems in that census tract equally among them.

The methods described above allowed us to assign a value for each census tract for specific drinking water contaminants: DBCP, nitrate, and arsenic. Based on the available data, we elected not to calculate any statistical measurers of precision, because the measurements were done with different analytic laboratory procedures. Our objective was to use only point estimates to characterize the exposures of census tracts.

Table 3 shows the frequency distribution of the DBCP estimates for the 109 census tracts in Fresno County. Since the City of Fresno represents over half of the County's population, the DBCP estimate grouping which includes the city is the largest. Fourteen (12.8 per cent) of the census tracts had DBCP estimates greater than 1 ppb. Surface water, not well water, is the primary source of drinking water for three of the four census tracts with no values for DBCP. The fourth census tract is a mountainous area in which DBCP was not used. The State Action Level for DBCP is 0.5 ppb while the maximum contaminant level is 1 ppb. These levels are not based on known or proven toxicologic impairments. For nitrate and arsenic, none of the census tracts with available data exceeded the MCLs. There was no correlation between DBCP levels and those of either arsenic ($r = 0.079$) or nitrates ($r = 0.119$).

TABLE 3—Frequency Distribution of DBCP Contamination Estimates by Census Tract

DBCP Contamination Estimate (ppb)	Number of Census Tracts	Percentage
None*	4	3.7
<0.01	4	3.7
0.01–0.19	60	55.0
0.01–0.19	10	9.2
0.2–0.49	13	11.9
0.5–0.99	4	3.7
≥1.0	14	12.8

N = 109.
*Surface water or no data.

Discussion

In the process of determining how water was distributed in Fresno County and how to calculate the water contamination levels, we had to make various assumptions. For water systems with more than one well, all were found to be open systems (i.e., water from one particular well would most likely supply those service connections in reasonable proximity to the well and not those service connections distant from the well). On the other hand, this could vary by the season of the year, quantity of water pumped from a specific well, etc. We found no reasonable method to determine what water went from any particular well to what particular service connection over time within a specific water system. After considerable discussion with governmental and water purveyor officials, we decided that we would have to assume that all water was mixed within the particular system and that each service connection would have to be treated equally. Thus, we used the arithmetic mean contaminant value for each water system adjusted for water production. In calculating an average, we elected not to transform the data, e.g., logarithmic, square root, etc. While this may be only an approximation of the actual value, there are no alternative methods available.

A few wells in some water systems were tested more frequently. In order not to have the mean value for the specific water system overrepresented by a minority of wells, we took the mean of each well's water contamination values before multiplying by its water production figures. We were unable to determine for the private wells if any particular well was tested more than once due to the restrictions concerning identity of the particular well. Of the three different sources for private well data, only in the 1500 wells in which owners requested the County conduct DBCP tests would one expect to find possible duplications. Again, based on conversations with County officials, such duplications would represent a small minority of the total wells.

We went to considerable lengths to obtain as much information as possible on both water systems and water quality. For example , we coded the well permit data for the County to assure ourselves that the exposure information would be as complete as possible.

The determination of the various contaminant levels by census tract was straightforward for those census tracts in which there was only one large water system. In the cases in which a large water system provided a portion of the census tract, we elected to assign the proportion supplied based on the geographic percentage of the census tract supplied by the water system, since we did not have data regarding the population density within various areas of any particular census tract. This underestimates the actual effect for the large system and overrepresents the effects from the small and private well systems.

We found little correlation between the DBCP census tract values and those for nitrates and arsenic. Both of the latter contaminants can, and usually do, occur naturally in the substrate, whereas DBCP is a totally manmade chemical and would only be present as a result of human activities. The heaviest DBCP census tracts in Fresno County correspond to those with the greatest potential (and actual use) for perennial crops, e.g., vineyards, orchards, etc. DBCP was especially effective for such crops in that the chemical effectively killed and controlled the nematodes but did not harm the plant.

Assessment of historical exposure is frequently one of the most difficult tasks in both environmental and occupational epidemiologic studies. Interpretation of study results is often severely limited due to inadequate or incomplete exposure assessment. We have described a practical methodology in developing such historical exposure data and believe that we have made as complete an exposure assessment as practically feasible, short of questioning each individual. Our effort to be as complete as possible in determining exposure information was both large and complex. We were fortunate to have the complete cooperation and assistance of various state, county, and municipal agencies as well as private water purveyors. Based on our exposure assessments, we believe that the interpretation of subsequent epidemiologic studies of water supply contaminants in Fresno County should not be hampered by inadequate exposure information although we recognize that many of our assumptions can be questioned. We also recognize that the water contamination data were primarily obtained from routine analyses and not subject to rigorous experimental design. Such is the quality of the data. Finally, we recognize that distribution of drinking water and consumption may not be similar. This is especially true if residents use commercially available bottled water.

For anyone having to collect similar data, we would offer these observations. Data from the large water systems were relatively easy to collect while those from the small systems were by far the hardest, requiring an enormous effort for data covering a relatively small percentage of the population. Water quality information on private wells was much more accessible because the county required such and collected data on any new well dug after 1977. Collecting private well water quality information from individual well owners would be prohibitive. Many assumptions must be made, e.g., average contamination in multiwell systems; accuracy of laboratory analyses; development of census tract means in census tracts with more than one water system; lack of DBCP contamination information prior to 1979, etc. While all of the assumptions make the results imprecise, they are the best estimation that can be feasibly and reasonably done.

ACKNOWLEDGMENTS

We extend our special thanks to the citizens of Fresno County, the California Department of Health Services, the Fresno County Department of Health, officials of the numerous water purveyors, Charles Ross, DO, and members of our peer review committee (Sir Richard Doll, Professor Brian MacMahon, and Professor T. W. Anderson). We also would like to thank our staff at Environmental Health Associates, particularly Diana Tesh and Carolyn Klassen.

REFERENCES

1. Jackson RJ, Greene CJ, Thomas JT, Murphy EL, Kaldor J: Literature Review on the Toxicological Aspects of DBCP and an Epidemiological Comparison of Patterns of DBCP Drinking Water Contamination with Mortality Rates from Selected Cancers in Fresno County, California, 1970–79. Sacramento: California Department of Health Services, 1982.
2. Whorton MD, Krauss RM, Marshall S, Milby TH: Infertility in male pesticide workers. Lancet 1977; 2:1259–1261.
3. Whorton MD, Foliart DE: Mutagenicity, Carcinogenicity, and Reproductive Effects of Dibromochloropropane (DBCP). Mutation Research 1983; 123:13–30.

NIMH Grants to Fund Research Evaluating Preventive Interventions

The Prevention Research Branch of the National Institute of Mental Health is soliciting grant applications for research that rigorously evaluates preventive interventions aimed at psychological disorders and dysfunctions. Of particular interest are studies aimed at:

- socio-emotional problems in high risk infants and their families (e.g., children of depressed mothers);
- early risk behaviors (e.g., aggression or poor peer relations) in preschool and early elementary school aged children;
- conduct disorder in school aged children; and
- anxiety depression, and suicide in adolescents and adults.

Preventive interventions are implemented in high risk populations who are either asymptomatic, or who exhibit early symptoms which do not yet meet diagnostic criteria for disorder. Preventive interventions precede the need for treatment, and result in the avoidance, interruption, reversal, or cessation of the pathogenic process leading to disorder or dysfunctions.

The Prevention Research Branch is also interested in research on interventions which result in the development, maintenance, or enhancement of protective psychosocial competencies, such as coping skills.

Interventions may focus on one or more risk factors, address one or more developmental periods, and be delivered in one or more settings.

For further information and consultation on the development of grant proposals, contact: Doreen Spilton Koretz, PhD, Assistant Chief, Prevention Research Branch, National Institute of Mental Health, Room 14C-02, 5600 Fishers Lane, Rockville, MD 20857. Tel: (301) 443-4283.